IT & C

ISSN 2821 - 8469, ISSN – L 2821 - 8469, Volumul 2, Numărul 1, Martie 2023

Inteligența artificială, o provocare esențială

Nicolae Sfetcu

Sfetcu, Nicolae (2023), Inteligența artificială, o provocare esențială, *IT & C*, **2**:1, 3-6, DOI: 10.58679/IT30677, https://www.internetmobile.ro/inteligenta-artificiala-o-provocare-esentiala/

Publicat online: 21.02.2023

Inteligența artificială, o provocare esențială

Nicolae Sfetcu
nicolae@sfetcu.com

Artificial intelligence, an essential challenge

Abstract

Artificial intelligence has progressed to the point where it is an essential component in almost all sectors of today's modern economy, with a significant impact on our private, social and political lives. It was founded on the assumption that human intelligence can be described so precisely that it can be made a machine to simulate it. This raises philosophical arguments about the mind and the ethics of creating artificial beings endowed with human-like intelligence. Artificial intelligences are a source of a whole new set of problems of explainability, accountability and trust.

Keywords: artificial intelligence

Rezumat

Inteligența artificială a progresat până la punctul în care este o componentă esențială în aproape toate sectoarele economiei moderne actuale, cu un impact semnificativ asupra vieții noastre private, sociale și politice. Ea a fost întemeiată pe presupunerea că inteligența umană poate fi descrisă atât de precis încât să poată fi făcută o mașină să o simuleze. Acest lucru ridică argumente filozofice despre minte și etica creării de ființe artificiale înzestrate cu inteligență asemănătoare omului. Inteligența artificială sunt o sursă a unui set complet nou de probleme de explicabilitate, responsabilitate și încredere.

Cuvinte cheie: inteligența artificială

IT & C, Volumul 2, Numărul 1, Martie 2023, pp. 3-6
ISSN 2821 - 8469, ISSN – L 2821 – 8469, DOI: 10.58679/IT30677
URL: https://www.internetmobile.ro/inteligenta-artificiala-o-provocare-esentiala/

Inteligența artificială a progresat până la punctul în care este o componentă esențială în aproape toate sectoarele economiei moderne actuale, cu un impact semnificativ asupra vieții noastre private, sociale și politice. Revenirea inteligenței artificiale în lumina reflectoarelor în ultimii ani este legată în principal de învățarea automată din date, care a făcut un salt înainte datorită apariției megadatelor. Practic, datele ajung la masa critică de exemple de învățat, algoritmii descoperă tipare predictive și tipare ascunse în date, iar arhitecturile de înaltă performanță reușesc să compenseze resursele de calcul și stocare necesare. Astfel, inteligența artificială a reușit să abordeze cu succes provocări deschise de lungă durată, cum ar fi înțelegerea și traducerea textelor și vorbirii, recunoașterea conținutului imaginilor și a videoclipurilor, rezumarea și abstractizarea, și alte sarcini despre care se crede că necesită inteligență.

Aplicațiile inteligenței artificiale includ motoare de căutare web avansate (de exemplu, Google), sisteme de recomandare (utilizate de YouTube, Amazon și Netflix), înțelegerea vorbirii umane (cum ar fi Siri și Alexa), mașini cu conducere autonomă (de exemplu, Tesla), luarea automată a deciziilor și concurarea la cel mai înalt nivel în sisteme de joc strategice (cum ar fi șah și go). Pe măsură ce mașinile devin din ce în ce mai capabile, sarcinile considerate a necesita „inteligență" sunt adesea eliminate din definiția inteligenței artificiale, un fenomen cunoscut sub numele de efectul IA. De exemplu, recunoașterea optică a caracterelor este adesea exclusă din lucrurile considerate a fi inteligență artificială, devenind o tehnologie de rutină.

Studiul raționamentului mecanic sau „formal" a început cu filozofii și matematicienii din antichitate. Studiul logicii matematice a condus direct la teoria calculului a lui Alan Turing, care sugera că o mașină, amestecând simboluri la fel de simple ca „0" și „1", ar putea simula orice act imaginabil de deducție matematică. Această înțelegere conform căreia computerele digitale pot simula orice proces de raționament formal este cunoscută sub numele de teza Church-Turing.

Cu aproximativ zece ani în urmă, s-a observat că unele modele de învățare cunoscute de mult, până acum ineficiente pentru sarcinile menționate, dacă sunt generalizate în mod corespunzător și instruite pe un set mare de date de exemplu, pot face un salt calitativ. Într-adevăr,

astfel de metode pot învăța, din pixelii imaginilor sau din cuvintele textelor de exemplu, „conceptele” generale care permit să recunoască și să clasifice cu precizie imagini noi și texte noi.

Etapa actuală de dezvoltare a inteligenței artificiale prezintă puncte forte și puncte slabe. Pe de o parte, capacitatea de învățare a modelelor de inteligența artificială crește, aducând progrese extraordinare în viziunea robotică și conducerea autonomă, în traducerea automată a textului și a vorbirii, în diagnosticul medical, în evaluarea riscurilor, în analiza predictivă. Pe de altă parte, decalajul cu alte aspecte ale inteligenței artificiale crește, în special, cu raționamentul și interacțiunea persoană-mașină, aspecte centrale pentru dezvoltarea unei inteligențe artificiale umane, etice și antropocentrice, care este punctul central al abordării europene. Opacitatea și natura cutiilor negre ale modelelor de inteligența artificială sunt în creștere, împreună cu riscul de a crea sisteme expuse la prejudecăți în datele de instruire, sisteme pe care nici chiar experții nu reușesc să le înțeleagă. Instrumentele lipsesc pentru a permite dezvoltatorilor de inteligența artificială să certifice fiabilitatea modelelor lor. Este crucial să se injecteze în tehnologiile de inteligența artificială valori etice ale corectitudinii (cum să se evite deciziile neloiale și discriminatorii), acuratețea (cum să se furnizeze informații fiabile), confidențialitatea (cum să se protejeze confidențialitatea persoanelor implicate) și transparența (cum să se modele și să se ia decizii inteligibile tuturor părților interesate). Această abordare de proiectare sensibilă la valoare, care încă nu a fost pe deplin operaționalizată, este puternic necesară pentru a stimula acceptarea socială pe scară largă a inteligenței artificiale, fără a-i inhiba puterea.

Mai mult, pe măsură ce apar sisteme sociotehnice din ce în ce mai complexe, formate din mulți oameni care interacționează și sisteme inteligente și autonome, inteligența artificială capătă o dimensiune societală importantă. O observație cheie este că o mulțime de indivizi inteligenți (asistată de instrumente de inteligența artificială) nu este neapărat o mulțime inteligentă. Dimpotrivă, poate fi proastă în multe cazuri, din cauza efectelor de rețea nedorite, neintenționate și a comportamentului agregat emergent. Exemplele abundă în societatea contemporană.

Inteligența artificială a fost întemeiată pe presupunerea că inteligența umană poate fi descrisă atât de precis încât să poată fi făcută o mașină să o simuleze. Acest lucru ridică argumente filozofice despre minte și etica creării de ființe artificiale înzestrate cu inteligență asemănătoare omului. Aceste probleme au fost explorate prin mit, ficțiune și filozofie încă din antichitate. Literatura științifico-fantastică și futurologia au sugerat, de asemenea, că, cu potențialul și puterea sa enormă, inteligența artificială poate deveni un risc existențial pentru umanitate. Interacțiunea

dintre alegerile individuale se poate desfăşura dramatic în provocările globale legate de inegalitatea economică, durabilitatea mediului și democrația. Efectele aglomerate ale rețelei și ale societății ale AI și impactul acestora asupra societății nu sunt suficient discutate în public și nu sunt abordate suficient de cercetarea inteligenței artificiale, în ciuda importanței izbitoare de a înțelege și de a prezice rezultatele agregate ale sistemelor bazate pe inteligența artificială sociotehnice şi ale proceselor sociale complexe conexe, precum și modalitatea de a evita efectele lor nocive. Astfel de efecte sunt o sursă a unui set complet nou de probleme de explicabilitate, responsabilitate și încredere, chiar presupunând că putem rezolva aceste probleme pentru un sistem individual de inteligența artificială bazat pe învățarea automată. Prin urmare, nu ne putem concentra exclusiv pe sensibilizarea cetățenilor sau instituțiilor individuale și capabile să ia decizii în cunoștință de cauză. De asemenea, trebuie să studiem efectele rețelei emergente ale mulțimilor de agenți interacționanți inteligenți, precum și proiectarea mecanismelor de colaborare distribuită care împing spre realizarea setului convenit de valori și obiective la nivel colectiv: mobilitate durabilă în oraşe, diversitate și pluralismul în dezbaterea publică și distribuția echitabilă a resurselor economice. Inteligența artificială socială se dezvoltă ca un nou domeniu de investigaţie cu impact potențial imens, necesitând următorul pas înainte în integrarea trans-disciplinară a inteligenței artificiale cu știința datelor, științele sociale, psihologia, știința rețelelor și sistemele complexe.